**International Journal of Academic and Applied Research (IJAAR)**
**ISSN: 2643-9603**
**Vol. 6 Issue 6, June - 2022, Pages: 118-124**

# Factors that Affecting Male Tuberculosis in East Java

**Diah Puspita Ningrum, Toha Saifudin, Suliyanto, Nur Chamidah**

Statistics Study Program, Departement of Mathematics, Airlangga University, Surabaya, Indonesia
Correspondence author : diah.puspita.ningrum-2018@fst.unair.ac.id, tohasaifudin@fst.unair.ac.id, yanfit@yahoo.com, nur-c@fst.unair.ac.id

***Abstract:*** *The infectious killer disease after* Coronavirus Disease 2019 *is tuberculosis. The existence of these conditions makes tuberculosis a world health issue. In 2020 tuberculosis cases in East Java were 42,922. Tuberculosis cases in East Java are dominated by males. The high case of tuberculosis in males is also felt in other countries such as Peru, Romania, and South Africa where cases of tuberculosis in males occur about 2 times higher than in females. Male tuberculosis cases in East Java were 11.2% higher than in females. This study was conducted to determine the factors that affect male tuberculosis in East Java. The results of the study using the Geographically Weighted Regression method showed that the factors that affect male tuberculosis in East Java are public places meeting health requirements, youth who smoke tobacco daily, and density population.*

**Keywords:** Tuberculosis, Male, East Java, Factor-Factor, Geographically Weighted Regression

## 1. INTRODUCTION

Tuberculosis is a disease that is transmitted by the bacterium *Mycobacterium tuberculosis*. Tuberculosis is included in the third goal of the 17 targets of the SDGs (*Sustainable Development Goals)*. The purpose of the SDGs is to ensure a healthy life and improve the welfare of all residents of all ages. Nowadays, tuberculosis is the second infectious killer disease after *Coronavirus Disease 2019* [9]. The existence of these conditions makes tuberculosis a world health issue, especially in Indonesia. In 2020, Indonesia became the largest contributor to tuberculosis sufferers and ranked 2nd out of 8 countries in the world after India. Tuberculosis cases in Indonesia in 2020 amounted to 845,000 and East Java is included in the province with high cases [7]. Tuberculosis cases in East Java were 42,922 and East Java was in 8th place [2]. Although in 2020 East Java ranks 8th, the success rate for treating TB cases in East Java has not met the target. The success rate of tuberculosis treatment in East Java in 2020 is 53%.

According to World Health Organization (WHO), gender affects the occurrence of tuberculosis. The male is more susceptible to infection with tuberculosis than women. In East Java, the incidence of male cases is higher than in females, as well as in other countries such as Peru, Romania, and South Africa where the incidence of TB in males is about 2 times higher than that of females [8]. Tuberculosis cases that occurred in East Java in males amount to 23,875 cases where male patients were 11.2% higher than women. Tuberculosis has a spatial dependence because tuberculosis is an infectious disease and is not limited to administrative areas. *The geographically Weighted Regression* (GWR) method can be used as a method to analyze spatial effects with a point approach. Using the GWR method can provide more information than the model formed.

Based on the facts, this study will discuss the factors that affect male tuberculosis in East Java with the predictor variables used are public places meeting health requirements, youth who smoke tobacco daily, male population, density population, and poor people. The results of this study are expected to be used as recommendations by the government in every district/city in East Java to tackle tuberculosis.

## 2. RESEARCH METHOD

### 2.1 Method and Data Source

The method used in determining the factors that affect male tuberculosis in East Java is Geographically Weighted Regression (GWR). The data used in this study are data obtained from the East Java Provincial Health Office and the Central Bureau of Statistics in East Java Province in 2020. The data used in this study include public places meeting health requirements, youth who smoke tobacco daily, male population, density population and poor people. The data obtained from the East Java Provincial Health Office are public places meeting health requirements and density population [2]. The data obtained from the Central Bureau of Statistics of East Java Province are youth who smoke tobacco daily, male population, and poor people [1].

### 2.2 Research Variable

This study used 38 units of observation in 38 district/city in East Java. Research variable that used in this study are presented in Table 1. below.

Table 1. Reaserach Variable

| Variables | Type of Variable | Definition |
|---|---|---|
| Male Tuberculosis in East Java (Y) | Numeric | The percentage of male tuberculosis Patients |
| Public Places Meeting Health Requirements ($X_1$) | Numeric | The percentage of public places meeting health requirements (%) |
| Youth who Smoke Tobacco Daily ($X_2$) | Numeric | The percentage of youth who smoke tobacco daily (%) |

| | | |
|---|---|---|
| Male Population ($X_3$) | Numeric | Amount of male population |
| Density Population ($X_4$) | Numeric | Density population per km$^2$ |
| Poor People ($X_5$) | Numeric | The percentage of poor people |

### 2.3 Step of the Research

a. Analyzing descriptive statistics for each research variable in the form of maximum, minimum, and average values. Research variables are also presented in thematic maps.
b. Performing a normality test on the data, testing the normality of the residuals using the Kolmogorov-Smirnov [6].

$$D = max_x|F^*(X) - S_n(X)| \quad (1)$$

where $S_n(X)$ is the cumulative distribution function of the sample and $F^*(X)$ is the cumulative normal distribution function.
c. Perform a multicollinearity test using the *Variance Inflator Factor* (VIF) value [4].

$$VIF_j = \frac{1}{1-R_j^2} \quad (2)$$

d. Testing the assumption of spatial dependence with Moran's I [5].

$$Z_I = \frac{I-E(I)}{\sqrt{Var(I)}} \quad (3)$$

Where $I$ *is* an index obtained from $I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}(y_i-\bar{y})(y_j-\bar{y})}{\sum_{i=1}^n (y_i-\bar{y})^2}$ and $E(I) = \frac{-1}{(n-1)}$
e. Testing the assumption of spatial heterogeneity with the Breusch Pagan [5].

$$BP = \left(\frac{1}{2}\right) \boldsymbol{f^t Z (Z^t Z)^{-1} Z^t f} \quad (4)$$

for $\boldsymbol{Z}$ matrices whose $(n \times (p+1))$ size and the vectors in it have been standardized (each observation) and $\boldsymbol{f}$ vector element is $f_i = (\frac{ei^2}{\hat{\sigma}^2} - 1)$ with $e_i = y_i - \hat{y}_i$
f. Determine the optimum bandwidth based on the *Cross Validation* (CV) [3].

$$CV(h) = \sum_{i=1}^n (y_i - \hat{y}_{\neq i}(h))^2 \quad (5)$$

With $\hat{y}_{\neq i}(h)$ value estimate from $y_i$ which the observations at each $(u_{i,}v_i)$ location are omitted in the estimation process.
g. Selecting the kernel weighting function to calculate the weighting matrix [3].

Kernel Gaussian function

$$w_j(u_{i,}v_i) = exp\left[-\frac{1}{2}\left(\frac{d_{ij}}{h}\right)^2\right] \quad (6)$$

And Kernel Bisquare function

$$w_j(u_{i,}v_i) = \begin{cases} \left[1 - \left(\frac{d_{ij}}{h}\right)^2\right]^2 & ,jika\ d_{ij} \leq h \\ 0 & ,jika\ d_{ij} > h \end{cases} \quad (7)$$

$h$ is bandwidth.
h. Estimating the GWR model [3].

$$\widehat{\boldsymbol{\beta}}(u_{i,}v_i) = \left(\boldsymbol{X^t W}(u_{i,}v_i)\boldsymbol{X}\right)^{-1}\boldsymbol{X^t W}(u_{i,}v_i)\boldsymbol{y} \quad (8)$$

where $\widehat{\boldsymbol{\beta}}(u_{i,}v_i)$ is the estimator of the $\boldsymbol{\beta}(u_{i,}v_i)$ parameters and $\boldsymbol{W}(u_{i,}v_i)$ is the weighting matrix.
i. Perform a conformity test on the GWR model, the hypothesis is:
$H_0$: $\beta_k(u_1, v_1) = \cdots = \beta_k(u_n, v_n) = \beta_k$
$H_1$:at least one $\beta_k(u_i, v_i) \neq \beta_k$ ; $k = 1,2, \dots, p, i = 1,2, \dots, n$
The test statistics [3]:

$$F_{hit} = \frac{(SSE(H_0) - SSE(H_1))/v}{SSE(H_1)/\delta_1} \quad (9)$$

where$SE(H_0) = \hat{\boldsymbol{\varepsilon}}^{\boldsymbol{t}}\hat{\boldsymbol{\varepsilon}} = (\boldsymbol{y} - \hat{\boldsymbol{y}})^{\boldsymbol{t}}(\boldsymbol{y} - \hat{\boldsymbol{y}}) = \boldsymbol{y^t}(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{y}$ with $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X^t X})^{-1}\boldsymbol{X^t}$
j. Perform a partial parameter significance test on the GWR model, the hypothesis:
$H_0$: $\beta_k(u_i, v_i) = 0$
$H_1$: $\beta_k(u_i, v_i) \neq 0; k = 1,2, \dots, p; i = 1,2, \dots, n$
The test statistics [3]:

$$T_{hit} = \frac{\hat{\beta}_k(u_i,v_i)}{\hat{\sigma}\sqrt{c_{kk}}} \quad (10)$$

with $c_{kk}$ elements from the diagonal of the k matrix $\boldsymbol{CC^t}$where $\boldsymbol{C} = \left(\boldsymbol{X^t W}(u_{i,}v_i)\boldsymbol{X}\right)^{-1}\boldsymbol{X^t W}(u_{i,}v_i)$.
k. Define the factors which affect male tuberculosis in every district/city in East Java.

## 3. RESULT AND DISCUSSION

### 3.1 Descriptive Statistics

Descriptive statistics analysis in this study is an analysis that contains maximum, minimum, and mean for each research variable. The analysis is presented in Table 2. below:

Table 2. Descriptive Statistics

| Variable | Maximum | Mean | Minimum |
|---|---|---|---|
| Y | 62.2 | 55.88 | 48.2 |
| $X_1$ | 96 | 68.07 | 18.8 |
| $X_2$ | 29.78 | 22.93 | 16.84 |
| $X_3$ | 1425168 | 533989.3 | 65685 |

| | | | |
|---|---|---|---|
| $X_4$ | 8286.5 | 1894.86 | 279.8 |
| $X_5$ | 22.78 | 11,02 | 3.89 |

Based on Table 2. the mean of variable Y is 55.88, the mean of variable x1 is 68.07, the mean of variable x1 is 22.93, the mean of variable x1 is 533989.3, the mean of variable x1 is 1894.86, and the mean of variable x1 is 11.02. Furthermore, each research variable will be presented in the form of a thematic map.

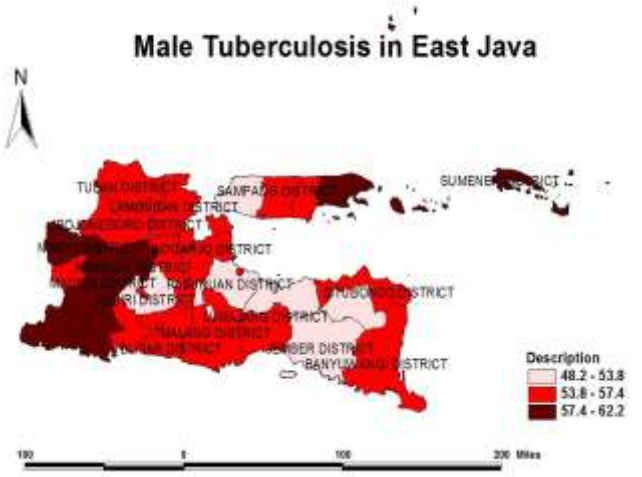


Figure 1 Male Tuberculosis in East Java

Based on Figure 1 distribution of male tuberculosis data from 38 districts/cities in East Java, the highest tuberculosis cases were in Madiun City and the lowest in Probolinggo District.

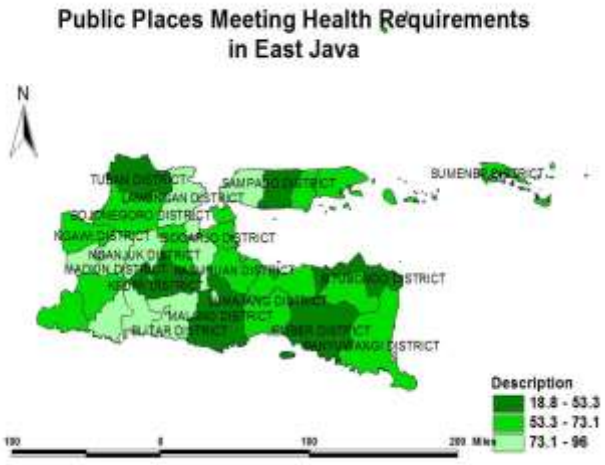


Figure 2 Public Places Meeting Health Requirements in East Java

Based on Figure 2, the distribution of public places meeting health requirements from 38 regencies/cities in East Java, the highest public places meeting health requirements is in Madiun City and the lowest is in Sampang District.

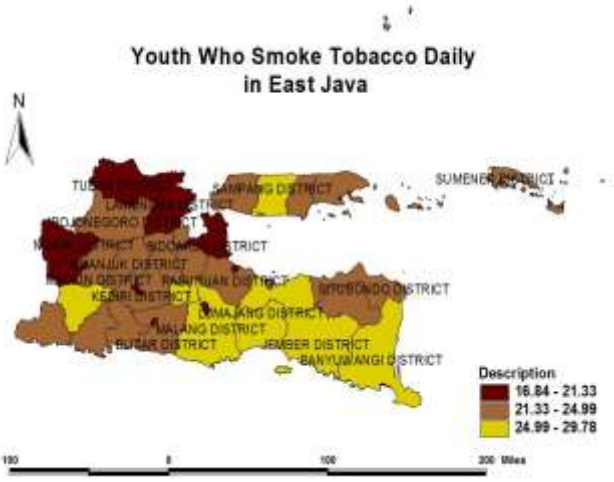


Figure 3 Youth Who Smoke Tobacco Daily in East Java

Based on Figure 3, the distribution of youth who smoked tobacco daily from 38 districts/cities in East Java, the highest number of youth who smoked tobacco daily was in Batu City and the lowest was in Surabaya City.

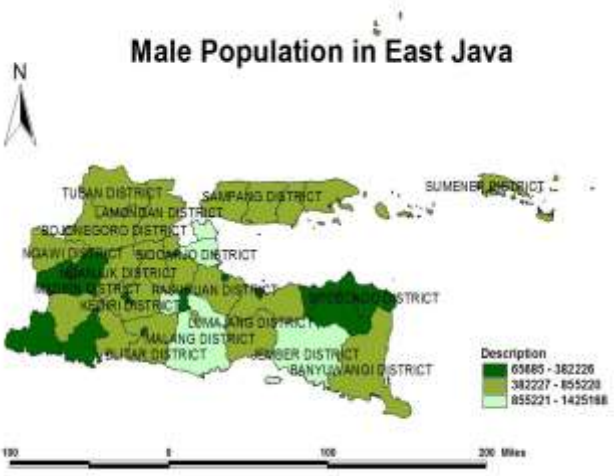


Figure 4 Male Population in East Java

Based on Figure 4, the distribution of male population data from 38 regencies/cities in East Java, the highest male population is in Surabaya City and the lowest is in Mojokerto City.

**International Journal of Academic and Applied Research (IJAAR)**
**ISSN: 2643-9603**
**Vol. 6 Issue 6, June - 2022, Pages: 118-124**

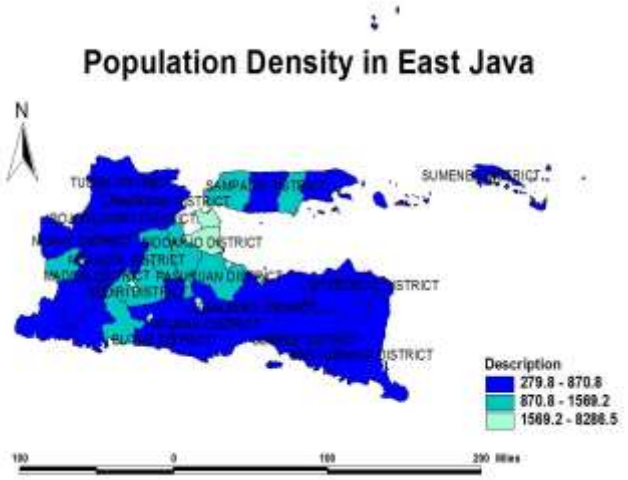


Figure 5 Population Density in East Java

Based on Figure 5 distribution of density population data from 38 regencies/cities in East Java, the highest density population is in Surabaya City and the lowest is in Banyuwangi District.

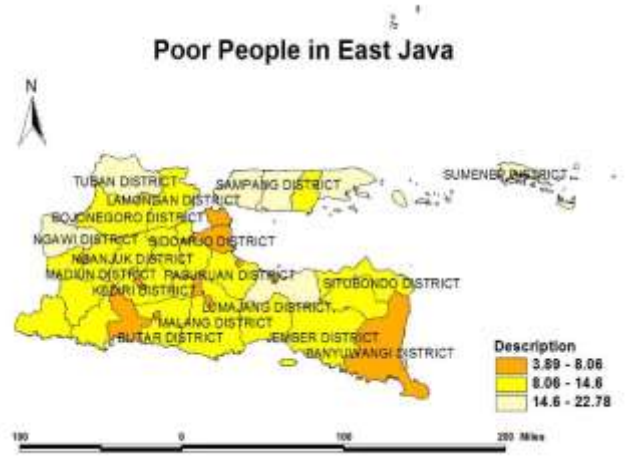


Figure 6 Poor People in East Java

Based on Figure 6 distribution of poor people data of 38 regencies/cities in East Java, the highest number of poor people are in Sampang District and the lowest are in Batu City

### 3.2 Normality Test and Multicollinearity Test

Normality test on male tuberculosis data using Kolmogorov Smirnov, the hypothesis is:

$H_0$: Residuals are normally distributed

$H_1$: Residuals are not normally distributed

The critical area in the *Kolmogorov-Smirnov test* is rejected if D > D(38; 0.1) = 0.130 or *p-value* < (10 %). With the help of *Software R* obtained a D value of 0.074 and *a p-value* of 0.87, then the decision is accepted so that it can be concluded that the residual male tuberculosis data in East Java is normally distributed. This shows that the data has met the assumption of normality. Furthermore, the multicollinearity test with the VIF value is in the Table 3.

Table 3. VIF Value

| VIF | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| Value | 1,97111 | 2.09882 | 1.27918 | 2.77922 | 2.29042 |

Based on Table 3. obtained a VIF value < 10 for each predictor variable, it can be concluded that there are no cases of multicollinearity, and the assumption of multicollinearity is met.

### 3.3 Spatial Dependency Test and Spatial Heterogneity Test

Modeling Tuberculosis using GWR, it is necessary to test the assumption of the spatial aspect with two test steps, namely the spatial dependence test and the spatial heterogeneity test. One statistical test that can be used to determine the presence of spatial dependence is Moran I. Spatial heterogeneity can be tested using Breusch Pagan. The result of Moran I test and Breusch Pagan test for male tuberculosis data are presented in Table 4.

Table 4. Moran I Test and Breusch Pagan Test Result

| Test | P-Value |
|---|---|
| Moran I test | 0.0000311 |
| Breusch Pagan test | 0.09254 |

Based on Table 4. P-Value of Moran I test and Breusch Pagan test are < $\alpha = 0.1$, then the decision is reject $H_0$, so there is spatial dependence and spatial heterogeneity in the data of male tuberculosis.

### 3.4 GWR Model

Tuberculosis modeling with the GWR approach was carried out using spatial weighting. The weighting matrix used in the GWR approach is a matrix containing kernel functions. The first step in GWR modeling is to determine the Euclidean distance between observation locations. The next step is to determine the best weighting function in TB modeling with the GWR approach where each weighting has a different bandwidth value. The chosen optimum weight determines the bandwidth value used. The optimum bandwidth value is obtained based on the minimum Cross-Validation (CV) criteria generated by the weighting function. The results of CV calculations are presented in Table 5. as follows:

Table 5. CV value for weighting function

| Function | CV |
|---|---|
| Fixed Gaussian Kernel | 8,001 |
| Fixed Bisquare Kernel | 7,994 |

In Table 5. shows that the fixed gaussian kernel has the minimum CV value, which is 8,001. The fixed gaussian kernel has a bandwidth of 0.591. The estimation model of male tuberculosis in East Java will be carried out using the GWR approach with fixed gaussian kernel weighting. The next step is to test the suitability of the GWR model. With the value of $\alpha = 10\%$, test criteria $H_0$ rejected if $F > F_{(\alpha;df_1;df_2)}$ $or$ $F > F_{(0.1;11.346;20.654)}$ $or$ $F > 1.8939$. By using software R, the results of the GWR model suitability test are presented in Table 6. as follows:

Table 6. The Reuslt of The GWR Model Suitability Test

| Source | DF | SS | MSE | F |
|---|---|---|---|---|
| GWR Improvement | 11.346 | 115.853 | 10.211 | 1.940 |
| GWR Residual | 20.654 | 108.710 | 5.263 | |

Table 5. shows that the F value is 1.940 > $F_{(0.1;11.346;20.654)}$, the decision taken is rejected $H_0$ so it can be concluded that the GWR spatial regression model obtained are suitable for the male tuberculosis in East Java.

After testing the suitability of the model, the next step is to test the significance of the GWR model parameters. The partial significance test of the GWR model parameters was carried out to determine which parameters had a significant effect on male tuberculosis in each district/city in East Java. The test statistic used to obtain predictor variables that significantly affect the response variable with a value of 10% is a $T$ test with the test criteria $H_0$ rejected if $|T_{ki}| > t_{(0,05;32)} = 1.69$. By using *software* GWR4 is obtained along with the coefficients of the predictor variables in the GWR model. For example, an analysis of the results of the estimation of model parameters is taken for the area that has the lowest percentage of male tuberculosis sufferers that is Probolinggo District. The hypotheses used to test the parameters in Probolinggo District are as follows:

$$H_0 : \beta_k(u_{13,}v_3) = 0; k = 1,2,3,4,5$$
$$H_1 : \beta_k(u_{13,}v_{13}) \neq 0; k = 1,2,3,4,5$$

The results of the partial test of the GWR model parameters are summarized in Table 7. as follows:

Table 7. The partial test of the GWR model parameters in Probolinggo District

| Parameter | Estimation | T Value | Decision |
|---|---|---|---|
| $\beta_0$ | 78.064618 | 9,074571 | - |
| $\beta_1$ | -0,071053 | -1,701154 | Reject $H_0$ |
| $\beta_2$ | 0,732519 | -3,187486 | Reject $H_0$ |
| $\beta_3$ | 0,0001 | -0,277333 | Accept $H_0$ |
| $\beta_4$ | -0,000546 | -1,493914 | Accept $H_0$ |
| $\beta_5$ | -0,106822 | -0,772966 | Accept $H_0$ |

Based on the partial test of the GWR model parameters in Probolinggo District in Table 7. the GWR model Probolinggo District is obtained as follows:

$$\hat{Y} = 78.064618 - -0,071053X_1 + 0,732519X_2$$

The following are the results of grouping variables that have a significant affect on male tuberculosis in each district/city in East Java which is presented in Table 8. as follows:

Table 8. Grouping of District/City in East Java Based on Predictor Variables with Significant Affect

| Group | District/City | Significant Variable |
|---|---|---|
| 1 | Pacitan District, Trenggalek District, Madiun District, Magetan District, Ngawi District, Madiun City | Public Places Meeting Health Requirements ($X_1$) |
| 2 | Ponorogo District, Tulungagung District, Bondowo District, Situbondo District, Pasuruan District, Gresik District, Sidoarjo District, Nganjuk District, Bojonegoro District, Tuban District, Lamongan District, Gresik District, Bangkalan | Youth who Smoke Tobacco Daily ($X_2$) |

| | | |
|---|---|---|
| | District, Sampang District, Pamekasan District, Sumenep District, Probolinggo City, Pasuruan City, Surabaya City | |
| 3 | Banyuwangi District | Density Population ($X_4$) |
| 4 | Probolinggo District | Public Places Meeting Health Requirements ($X_1$) and Youth who Smoke Tobacco Daily ($X_2$) |
| 5 | Blitar District, Kediri District, Malang District, Mojokerto District, Jombang District, Kediri City, Blitar City, Malang City, Mojokerto City, Batu City | Youth who Smoke Tobacco Daily ($X_2$) and Density Population ($X_4$) |
| 6 | Lumajang District | Public Places Meeting Health Requirements ($X_1$), Youth who Smoke Tobacco Daily ($X_2$), and Density Population ($X_4$) |

Based on Table 8. it can be seen that factors that have a significant affect on male tuberculosis in East Java are public places meeting health requirements, youth who smoke tobacco daily, and density population. The thematic map of the factors that have a significant affect by grouping can be seen in Figure 7 below:

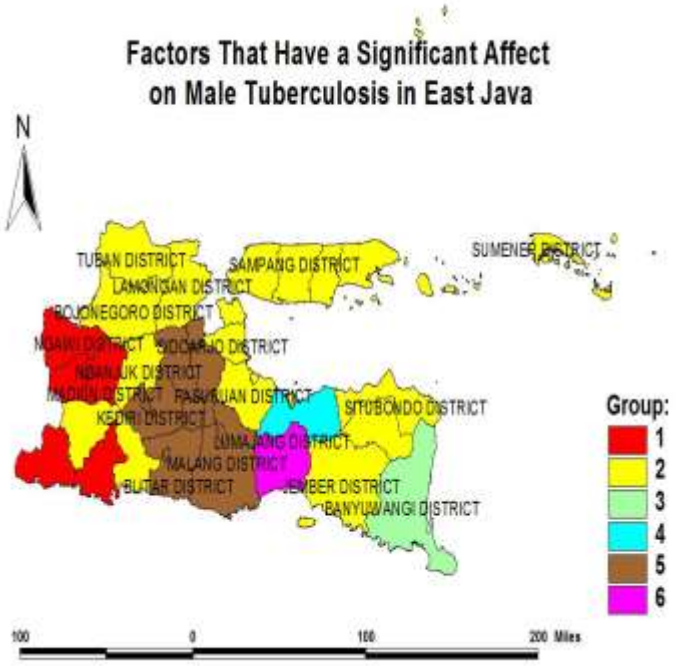


Figure 7 Thematic Map of Factors That Have a Significant Affect on Male Tuberculosis in East Java

Based on Figure 7. showing the distribution of predictor variables that have a significant effect, there are 6 district/city groupings which show the similarity of factors that significantly affect male tuberculosis in each district/city in East Java spatially.

## 4. CONCLUSION

The distribution of male tuberculosis data from 38 districts/cities in East Java, the highest tuberculosis cases were in Madiun City and the lowest case were in Probolinggo District. The results of the study using the Geographically Weighted Regression method showed that the factors that affect male tuberculosis in East Java are public places meeting health requirements, youth who smoke tobacco daily, and density population. From the distribution of predictor variables that have a significant effect, there are 6 groupings of districts/cities which show the similarity of factors that significantly affect male tuberculosis in each district/city in East Java.

Based on this study, the government in each district/city in East Java can make policies to improve the factors that influence male tuberculosis. Especially in Madiun City, which is the area with the highest number of male tuberculosis sufferers. Not only the government, but the role of the community is also needed in suppressing tuberculosis cases in East Java.

## 5. REFERENCES

[1] BPS. 2020. *Indikator Tujuan Pembanguan Berkelanjutan Indonesia* 2020. Badan Pusat Statistik, Jakarta.

[2] Dinas Kesehatan Provinsi Jawa Timur. 2021. *Profil Kesehatan Jawa Timur 2020*. Dinas Kesehatan Provinsi Jawa Timur, Surabaya.

[3] Fotheringham, A. S., Brunsdon, C. dan Charlton, M. E., 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationship*. John Wiley & Sons Inc., New York.

[4] Gujarati, D. N., 2004. *Basic Econometrics*, 4th Edition. McGraw-Hill., New York.

[5] Lee, J. & Wong, D. W. S., 2001. *Statistical Analysis with Arcview GIS.* John Wiley & Sons Inc., New York.

[6] Razali, N. M. dan Wah, Y. B. 2011. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*. Vol 2, No 1, 21-33.

[7] TB Indonesia. 2020. *Dashboard TB.* Kementerian Kesehatan Republik Indonesia. https://tbindonesia.or.id/pustaka-tbc/dashboard-tb/. [14 Januari 2022]

[8] Ugarte-Gill, C., Alisjahbana, B., Ronacher, K., Riza, A. L., Koesoemadinata, R. C., Malherbe, S, T., Ciobota, R., Llontop, J. C., Kleynhans, L., Lopez, S., Santoso, P., Marius, C., Villaizan, K., Ruslami, R., Walzl, G., Panduru, N. M., Dpckrell, H. M., Hill, P. C., Allister, S. Mc., Pearson, F., Moore, D. A. J., Critchley, J. A., Crevel, R. V., 2019. Diabetes Mellitus Among Pulmonary Tuberculosis Patients From 4 Tuberculosis-endemic Countries: The TANDEM Study. *Clinical Infectious Diseases*. Vol 70, No 5, 780-788.

[9] WHO. (2021). *Tuberculosis*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/tuberculosis. [20 Januari 2022]